Pitch perception and production in congenital amusia: evidence from Cantonese speakers

Article

Accepted Version

Liu, F., Chan, A. H. D., Ciocca, V., Roquet, C., Peretz, I. and Wong, P. C. M. (2016) Pitch perception and production in congenital amusia: evidence from Cantonese speakers. Journal of the Acoustical Society of America, 140 (1). pp. 563-575. ISSN 0001-4966 doi: https://doi.org/10.1121/1.4955182 Available at https://centaur.reading.ac.uk/66404/

To link to this article DOI: http://dx.doi.org/10.1121/1.4955182

Publisher: Acoustical Society of America

Pitch perception and production in congenital amusia: Evidence from Cantonese speakers

Fang Liu [b)]
School of Psychology and Clinical Language Sciences, University of Reading, Earley Gate, Reading RG6 6AL, UK

Alice H. D. Chan [a) b)]
Division of Linguistics and Multilingual Studies, School of Humanities and Social Sciences, Nanyang Technological University, S637332, Singapore

Valter Ciocca
School of Audiology and Speech Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Catherine Roquet and Isabelle Peretz
International Laboratory for Brain, Music and Sound Research (BRAMS), Université de Montréal, Montreal, Quebec, Canada

Patrick C. M. Wong[a)]
Department of Linguistics and Modern Languages and Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong, China

Running title: Congenital amusia in Cantonese speakers

Date submitted: June 14, 2016

---

a) Author to whom correspondence should be addressed. Electronic mail: p.wong@cuhk.edu.hk or alice@ntu.edu.sg. Also at The Chinese University of Hong Kong – Utrecth University Joint Center for Language, Mind and Brain or Division of Linguistics and Multilingual Studies at Nanyang Technological University.
b) Fang Liu and Alice H. D. Chan contributed equally to this work.

Abstract This study investigated pitch perception and production in speech and music in individuals with congenital amusia (a disorder of musical pitch processing) who are native speakers of Cantonese, a tone language with a highly complex tonal system. Sixteen Cantonese-speaking congenital amusics and 16 controls performed a set of lexical tone perception, production, singing, and psychophysical pitch threshold tasks. Their tone production accuracy and singing proficiency were subsequently judged by independent listeners, and subjected to acoustic analyses. Relative to controls, amusics showed impaired discrimination of lexical tones in both speech and non-speech conditions. They also received lower ratings for singing proficiency, producing larger pitch interval deviations and making more pitch interval errors compared to controls. Demonstrating higher pitch direction identification thresholds than controls for both speech syllables and piano tones, amusics nevertheless produced native lexical tones with comparable pitch heights/contours and intelligibility as controls. Significant correlations were found between pitch threshold and lexical tone perception, music perception and production, but not between lexical tone perception and production for amusics. These findings provide further evidence that congenital amusia is domain-general language-independent pitch-processing deficit that is associated with severely impaired music perception and production, mildly impaired speech perception, and largely intact speech production.

PACS numbers: 43.70.Mn, 43.75.Cd, 43.75.Rs

I. INTRODUCTION

Music and language are two defining aspects of human cognition. Musicianship and tone-language expertise (e.g., Mandarin, Cantonese) have been shown to assert beneficial effects on pitch processing in music and speech in a bidirectional way, with musicians showing superior processing of lexical tone (e.g., Bidelman et al., 2011; Lee and Hung, 2008; Wong et al., 2007) and tone-language speakers demonstrating enhanced processing of musical pitch (e.g., Bidelman et al., 2013; Deutsch et al., 2009; Pfordresher and Brown, 2009; Wong et al., 2012; but see Peretz et al., 2011 for negative support). A closely related proposal that has been studied extensively in recent years is that tone-language expertise would reduce or compensate for deficits in musical disorders such as congenital amusia ("amusia", hereafter; Peretz, 2008; Peretz et al., 2002). Evidence predominantly drawn from speakers of Mandarin, a tone language with four lexical tones (e.g., Jiang et al., 2010; Liu et al., 2012a; Nan et al., 2010), does not seem to support this proposal. However, there are tone languages in the world with more complex tonal systems than Mandarin (Yip, 2002). There are six lexical tones in Cantonese, named Tone 1 to Tone 6. Using the scale of 1-5 (5 being the highest and 1 being the lowest pitch level) (Chao, 1930), the high-level Cantonese Tone 1 can be transcribed as having pitch values of 55, the high-rising Tone 2 as having 25, the mid-level Tone 3 as 33, the low-falling Tone 4 as 21, the low-rising Tone 5 as 23, and the low-level Tone 6 as 22. The fact that there are multiple tones with similar shapes in Cantonese (Rising: 25/23; Level: 55/33/22; Falling: 21) makes tone perception challenging even for native speakers (Ciocca and Lui, 2003; Francis and Ciocca, 2003; Varley and So, 1995). It has been shown that Cantonese speakers demonstrate enhanced pitch processing abilities compared to Mandarin speakers (Zheng et al., 2010, 2014). Thus, it is possible that Cantonese speakers with amusia may exhibit different characteristics in pitch processing in speech and music than Mandarin speakers with amusia. The present study was conducted to explore this possibility.

Congenital amusia is a heritable neurodevelopmental disorder of musical processing (Drayna et al., 2001; Peretz et al., 2007), affecting around 3-5% of the general population across speakers of tone and non-tonal (e.g., English, French) languages (Kalmus and Fry, 1980; Nan et al., 2010; Peretz et al., 2008; Wong et al., 2012; although see Henry and McAuley, 2010, 2013 for contrasting views). Since the first systematic study of amusia published in 2002 (Ayotte et al., 2002), the majority of amusia studies have focused on speakers of non-tonal languages such as English (e.g., Foxton et al., 2004; Jones et al., 2009b; Liu et al., 2010; Loui et al., 2008) and French (e.g., Albouy et al., 2013; Dalla Bella et al., 2009; Hutchins et al., 2010a; Tillmann et al., 2011b). The first studies of amusia on tone language (Mandarin) speakers were published in 2010 (Jiang et al., 2010; Nan et al., 2010), and were followed by a series of studies on Mandarin speakers (e.g., Jiang et al., 2012b; Liu et al., 2012a, 2015a) and only two studies on Cantonese speakers (Liu et al., 2015b; Wong et al., 2012).

Across speakers of different language backgrounds (English, French, or Mandarin), individuals with amusia ("amusics", hereafter) have been reported to have severely impaired music perception and production (e.g., Ayotte et al., 2002), mildly impaired speech perception (e.g., Liu et al., 2010), and largely intact speech production (e.g., Nan et al., 2010). Amusics also demonstrated impaired fine-grained pitch discrimination, insensitivity to pitch direction (up versus down), and/or lack of pitch awareness for speech and music (Foxton et al., 2004; Hutchins and Peretz, 2012; Hyde and Peretz, 2004; Jiang et al., 2011; Liu et al., 2010, 2012a, 2012b, Loui et al., 2008, 2009, 2011, Moreau et al., 2009, 2013, Peretz et al., 2002, 2009; Tillmann et al., 2011b). Neuroimaging studies suggested reduced connectivity along the right frontotemporal pathway in the amusic brain for English/French speakers (Hyde et al., 2006, 2007, 2011; Loui et al., 2009).

Although primarily a musical disorder, amusia also impacts many aspects of speech processing for speakers of English, French, and Mandarin: First, amusics demonstrate impaired perception of intonation and of lexical tones when the pitch contrasts involved are relatively small (Hutchins et al., 2010a; Jiang et al., 2010, 2012a, 2012b, Liu et al., 2010, 2012a); second, amusics show impaired phonemic awareness for speech segments and lexical tones (Jones et al., 2009a; Nan et al., 2010); third, amusics show impaired discrimination of non-native lexical tones (Tillmann et al., 2011a); fourth, amusics demonstrate impaired ability to identify certain emotions in spoken utterances (Thompson et al., 2012); finally, amusics show impaired speech comprehension in both quiet and noisy background, and with either normal or flat F_0 (the fundamental frequency, in Hz) (Liu et al., 2015a).

While the aforementioned speech processing deficits in amusia are all related to perception, research has indicated that amusics' speech production abilities are largely preserved. In particular, English-speaking amusics showed normal laryngeal control (the regularity of vocal fold vibration) and pitch production (overall pitch range and pitch regularity) when reading a story and producing three sustained vowels (Liu et al., 2010). French-speaking amusics exhibited automatic vocal pitch shift in response to frequencyshifted auditory feedback when speaking and singing (Hutchins and Peretz, 2013). Furthermore, although Mandarin-speaking amusics demonstrated impaired lexical tone identification and discrimination, their production of lexical tones achieved near-perfect recognition rates by native listeners (Nan et al., 2010). Speech perception-production mismatch in amusia was further supported by the finding that amusics of English or French background demonstrated better imitation than identification or discrimination of speech intonation (Hutchins and Peretz, 2012; Liu et al., 2010). However, despite showing better production than perception of pitch in speech and music, amusics still do not perform at the normal level on pitch/timing matching when imitating speech/music materials. For example, Mandarin-speaking amusics showed impaired pitch and rhythm matching in speech and song imitation (Liu et al., 2013), which was consistent with the reduced performance on pitch matching of single tones and pitch direction in English- and French-speaking amusics (Hutchins et al., 2010b; Loui et al., 2008).

As reviewed above, increasing evidence has suggested that amusia is a domain-general pitch-processing deficit, regardless of language background (English, French, or Mandarin). The most commonly used diagnosis test for amusia is the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz et al., 2003), which consists of six subtests measuring the perception of scale, contour, interval, rhythm, and meter of Western melodies, and recognition memory of these melodies. Those who score two standard deviations below the control mean on the global score or scores on the first three pitch-based subtests are classified as amusics (Liu et al., 2010; Peretz et al., 2003). Thus, amusia is primarily a music perception disorder according to its diagnosis criteria. The majority of amusics also have impaired music production, being unable to sing in tune or imitate song accurately (Ayotte et al., 2002; Dalla Bella et al., 2009; Liu et al., 2013; Tremblay-Champoux et al., 2010).

It is a matter of debate whether speakers of tone languages have enhanced perception of pitch compared to non-tonal language speakers, with some studies showing positive evidence (Giuliano et al., 2011; Pfordresher and Brown, 2009) while other studies revealing negative results (Bent et al., 2006; Bidelman et al., 2011b; DiCanio, 2012; Schellenberg and Trehub, 2008; So and Best, 2010; Stagray and Downs, 1993). A meta-analysis of the amusia literature indicates that tone language experience has no impact on pitch processing in amusia (Vuvan et al., 2015). However, it has been shown that the cut-off score for the diagnosis of amusia is higher in Hong Kong for Cantonese speakers (78.4%) than in Canada for French speakers (73.9%), based on the On-Line Identification Test of Congenital Amusia (Peretz et al., 2008; Wong et al., 2012). This is in contrast to Mandarin speakers, who show a lower cut-off score (71.7%) than French speakers (74.7%) for the diagnosis of amusia using the MBEA global score (Nan et al., 2010). As mentioned earlier, Cantonese is a tone language with a greater degree of complexity in terms of tonal contrasts than Mandarin (Yip, 2002). Presumably because of being exposed to such a rich inventory of tonal patterns in everyday life, Cantonese speakers demonstrate comparable performance to English musicians on pitch and music processing (Bidelman et al., 2013), and enhanced categorical perception of pitch relative to Mandarin speakers (Zheng et al., 2010, 2014). Furthermore, compared with English- and French-speaking amusics, fewer Cantonese-speaking amusics reported to have problems with singing (Peretz et al., 2008; Wong et al., 2012). Our recent study also revealed intact brainstem encoding of speech and musical stimuli in Cantonese-speaking amusics (Liu et al., 2015b). Because of the exposure to a complex lexical tone system, Cantonese-speaking amusics ("Cantonese amusics", hereafter) may have developed strategies for compensating for pitch-processing deficits in speech and music. The present study investigated this issue using a comprehensive set of tasks, including lexical tone perception and production, singing, and psychophysical pitch thresholds.

In the lexical tone discrimination tasks, we used both speech and non-speech stimuli, and employed tone pairs with either subtle (tone pairs 21-22, 22-33) or relatively large (tone pairs 23-25, 25-55) pitch differences. In the lexical tone production task, Cantonese amusics and controls produced six lexical tones on the same syllable *[si]* to represent different word meanings ("poem", "history", "exam", "time", "market", and "right"). An independent group of Cantonese listeners then identified the words that amusics and controls produced. In the singing task, Cantonese amusics and controls sang "*Happy birthday*" in Cantonese. Singing proficiency was evaluated using both acoustic analyses and subjective ratings by an independent group of native listeners. In the pitch threshold tasks, pitch direction identification thresholds were assessed using adaptive tracking procedures for speech syllables and piano tones.

II. METHOD

A. Participants

Participants (n = 16 in each group) were recruited by advertisements through mass mail services at the Chinese University of Hong Kong. All participants had normal hearing in both ears, with pure-tone air conduction thresholds of 25 dB HL or better at frequencies of 0.5, 1, 2, and 4 kHz. All were native speakers of Cantonese, and none reported having speech or hearing disorders or neurological/psychiatric impairments in the questionnaires concerning their music, language, and medical background. Written informed consents were obtained from all participants prior to the experiments. The Institutional Review Board of Northwestern University and The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee approved the study.

All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). The diagnosis of amusia in these participants was conducted using the MBEA (Peretz et al., 2003). Those who scored 65 or under on the pitch composite score (sum of the scores on the scale, contour, and interval subtests) or below 78% correct on the MBEA global score were classified as amusic, as these scores correspond to 2 standard deviations below the mean scores of normal controls (Liu et al., 2010; Peretz et al., 2003). Controls were chosen to match with amusics on sex, handedness, age, years of education, and musical training background, but having significantly higher MBEA scores (Table 1).

[Insert Table 1 about here]

B. Tasks

The tone perception, production, and singing tasks were administered to all participants in fixed order: 1) lexical tone perception (speech and non-speech in counterbalanced order across participants), 2) lexical tone production, and 3) singing of "*Happy birthday*" in Cantonese (which has the same melody as the English version). These three tasks took around 30 minutes on average. Depending on their availability, participants also participated in two pitch threshold tasks on a different day. All experiments were conducted in a soundproof booth at the Chinese University of Hong

Kong. The perception tasks were delivered through Sennheiser HD 380 PRO Headphones at a comfortable listening level. The production data were recorded directly onto a desktop PC using Praat (Boersma and Weenink, 2001) with a Shure SM10A headworn microphone and a Roland UA-55 Quad-Capture audio interface, at a sampling rate of 44.1 kHz.

1. Lexical tone perception

To assess subjects' perceptual ability on lexical tones, four Cantonese tone pairs with varying degrees of tonal contrasts were used in the current tone perception tasks for both speech (Cantonese syllables) and non-speech (low-pass filtered hums; filtered at 1900 Hz) conditions. Fig. 1 shows time-normalized F_0 contours of these tone pairs. These stimuli were a subset of stimuli from Wong et al. (2009), chosen based on varying degrees of tonal contrasts for the current purpose, i.e., investigating amusics' tone perception deficits in subtle tone pairs (tone pairs 21-22, 22-33; Fig. 1A and 1C), while including tone pairs that have relatively coarse differences for comparison (tone pairs 23-25, 25-55; Fig. 1B and 1D). Normalized in amplitude and duration, the stimuli in each pair only differed in F_0 characteristics. It is worth noting that F_0 is the dominant cue that listeners use in tone perception, whereas the contribution of amplitude and duration is only secondary (Abramson, 1962; Lin and Repp, 1989; Liu and Samuel, 2004; Tong et al., 2015; Whalen and Xu, 1992). Thus, our normalization of amplitude and duration in these tone pairs is unlikely to have had significant impact on discrimination of these tones.

The difference in mean F_0 (averaged F_0 value across the tone) between bei22 (low-level, Tone 6) and bei33 (mid-level, Tone 3) was 2.02 semitones (Fig. 1A). The difference in final F_0 (F_0 value at the tone offset) between jyu23 (low-rising, Tone 5) and jyu25 (high-rising, Tone 2) was 2.90 semitones (Fig. 1B), and that between min21 (low-falling, Tone 4) and min22 (low-level, Tone 6) was 1.92 semitones (Fig. 1C). The difference in initial F_0 (F_0 value at the tone onset) between tong25 (high-rising, Tone 2) and tong55 (high-level, Tone 1) was 5.66 semitones, and the F_0 range (maximum F_0 – minimum F_0) between these two tones was 4.57 semitones (Fig. 1D). Therefore, the F_0 differences between these tone pairs are likely to result in different degrees of subtlety/saliency of pitch contrasts, with 21-22 and 22-33 having smaller pitch differences than 23-25 and 25-55.

[Insert Fig. 1 about here]

In the lexical tone perception task, participants were required to discriminate the tones/pitches of the four Cantonese word pairs in speech and non-speech contexts, blocked by context. The order of presentation of speech and non-speech contexts was counterbalanced across participants. Four practice trials (each with 5 repetitions; with stimuli different from those in experimental trials) were given to the participants to familiarize them with the procedure and stimuli. In total, 80 stimulus pairs (4 tone pairs × 2 "same" sequences × 2 "different" sequences × 5 repetitions) were presented to the participants in pseudorandom order in each condition (speech versus non-speech), with the inter-stimulus interval of 500 milliseconds. The participants' task was to click the appropriate button on the computer screen to indicate whether the tones/pitches of the two stimuli in each pair were the same or different.

2. Lexical tone production

To examine possible mismatches between tone perception and production, participants were asked to read aloud sentences including all six tones on the same syllable *[si]*, as in 詩 */si55/*, 史 */si25/*, 試 */si33/*, 時 */si21/*, 市 */si23/*, and 是 */si22/*. These tones were embedded in six Chinese words, 詩歌 ["poem"], 歷史 ["history"], 嘗試 ["exam"], 時間 ["time"], 街市 ["market"], and 是非 ["right and wrong"] for the participant's reference. The task required participants to read aloud the target syllables in a carrier sentence "下一個字係: __" ["The next word is: __"]. Participants were required to produce the target tones only (i.e., 詩, 史, 試, 時, 市, and 是), and not both syllables in each word. Each sentence was produced three times at a normal pace and in random order.

3. Singing

In the singing task, participants were asked to sing the "*Happy birthday*" song in Cantonese at a normal pace while their voice was recorded.

4. Pitch threshold tasks

To further examine pitch perception of amusics, all but 3 participants (1 amusic and 2 controls did not participate due to unavailability) also completed two psychophysical pitch threshold tasks, during which they were required to identify the direction of pitch movement (high-low versus low-high) in the speech syllable /ma/ and its piano tone analog. For both stimulus types, there were a standard stimulus of 131 Hz (C3) and 63 target stimuli that deviated from the standard in steps (ΔF, F_0 difference or pitch interval between the standard and target stimuli) of 0.01 (10 steps between 131.08 and 131.76 Hz, increasing by 0.01 semitones in each step), 0.1 (9 steps between 131.76 and 138.79 Hz, increasing by 0.1 semitones in each step), and 0.25 semitones (44 steps between 138.79 and 262 Hz, increasing by 0.25 semitones in each step). Thus, the smallest pitch interval (ΔF between the standard and step 1 deviant) between the standard and target stimuli was 0.01 semitones, and the largest pitch interval (ΔF between the standard and step 63 deviant) was 12 semitones.

Stimuli were presented with adaptive tracking procedures using the APEX 3 program developed at ExpORL (Francart et al., 2008). As a test platform for auditory psychophysical experiments, APEX 3 enables the user to specify custom stimuli and procedures with eXtensible Markup Language (XML). The “two-down, one-up” staircase method was used in the adaptive tracking procedure (Levitt, 1971), with step sizes of 0.01, 0.1, and 0.25 semitones as explained earlier. Following a response, the next trial was played 750 ms later. In the staircase, a reversal was defined as a change of direction, e.g., from “down” to “up”, or from “up” to “down”. Each run ended after 14 such reversals, and the threshold (in semitones) was calculated as the mean of the pitch intervals (pitch differences between the standard and target stimuli) in the last 6 reversals. It took on average around 15 minutes to complete the two pitch threshold tasks.

C. Tone production judgments and singing ratings

In order to assess whether and to what extent the tones produced by the amusic and control participants can be correctly recognized by native listeners, an independent group of eight Cantonese-speaking informants were asked to identify the words that were associated with the produced tones based on their own judgments. The eight listeners heard the tones produced by the 32 amusic/control speakers (576 in total, = 32 participants × 6 tones × 3 tokens), as in *section 2.2.2*, in random order. Their task was to identify the word they just heard by pressing the buttons representing the word 詩 (as in 詩歌), 史 (as in 歷史), 試 (as in 嘗試), 時 (as in 時間), 市 (as in 街市), and 是 (as in 是非), respectively. Their identification accuracy (in percentage of correct responses to the intended tones) and confusion matrices between the intended tones and their responses were calculated.

In order to assess singing proficiency of the amusic and control participants, the same eight Cantonese-speaking informants also participated in a singing-rating task, which consisted of two sessions. In the first session, listeners were required to rate whether the songs (presented in random order across listeners, with the group status unknown to the listeners) they heard were in-tune or not on a Likert scale of 1 (very out-of-tune) to 8 (very in-tune) based on their intuition. In the second session, listeners were required to rate the singing using an accuracy rating scale (1-8) developed by Wise and Sloboda (Anderson et al., 2012; Wise and Sloboda, 2008). The use of both intuition (session 1) and guidelines (session 2) for singing rating enabled comparisons of different rating criteria and the evaluation of the reliability of singing proficiency judgments.

D. Scoring and data analyses

Statistical analyses were conducted using R (R Core Team, 2014). Participants' performance on the tone discrimination tasks (in speech and non-speech) was scored using *d′* from signal detection theory (Macmillan and Creelman, 2005). The statistic *d′* was calculated using Table A5.4 for the differencing model in Macmillan and Creelman (2005). Hit (a "different" pair was judged as different) and false alarm (a "same" pair was judged as different) rates were adjusted up or down by 0.01 when their values were 0 or 1. Tone production accuracy of amusics and controls was scored using the percentage of correct responses by the 8 native listeners, transformed using rationalized arcsine transformation (Studebaker, 1985). In order to examine F_0 contours of the produced tones by each participant, acoustic analyses were conducted using the Praat script ProsodyPro (Xu, 2013). Ten F_0 data points (time-normalized) were extracted for each tone.

Besides singing ratings by the eight Cantonese listeners, acoustic analyses were done on each song using Praat. Among the 32 participants, 29 produced 25 notes and 3 produced 24 notes for the song. Each note's F_0 height in Hertz (Hz) and semitones was measured, as well as its rhyme duration in seconds. Adapting the acoustic measurements in previous singing studies (Dalla Bella et al., 2007, 2009), the following F_0 and time variables were calculated to examine singing proficiency of the participants: 1) *Pitch interval deviation* (in semitones, or st): the absolute difference between the produced interval (pitch distance between two consecutive notes; in absolute value) and the notated interval as determined by the musical score of the song. The bigger the value, the less accurate the singing in relative pitch. 2) *Number of pitch interval errors*: the number of produced intervals that deviated from the notated intervals by more than ±0.5 semitones. 3) *Number of contour errors*: the number of produced intervals that constituted different pitch directions (up, down, or level) than the notated intervals. 4) *Number of time errors*: the number of produced notes that were at least 25% longer or shorter than the notated durations.

III. RESULTS

A. Lexical tone perception

Fig. 2 shows performance of the two groups in the tone perception tasks under speech and non-speech conditions, with scores on the four different tone pairs plotted separately. A repeated-measures ANOVA on d' scores was conducted for each condition, with group (amusic, control) as the between-subject factor, tone pair (22-33, 23-25, 21-22, 25-55) as the within-subject factor, and subject as the random factor.

[Insert Fig. 2 about here]

For the speech condition, a significant main effect of group [$F(1,30) = 7.33$, $p = .011$] was observed, as controls achieved better performance than amusics. Performance also varied significantly across different tone pairs [$F(3,90) = 21.80$, $p < .001$]. Post-hoc pairwise *t*-tests (two-sided) with the Holm correction (Holm, 1979) indicated that discrimination of tone pair 25-55 was significantly better than that of other tone pairs (all *p*s < .05), and discrimination of tone pairs 21-22 and 22-33 was significantly better than that of 23-25 (both *p*s < .05). There was no significant group × tone pair interaction [$F(3,90) = 1.90$, $p = .135$].

For the non-speech condition, there was also a significant group effect [$F(1,30) = 4.19$, $p = .0495$], with controls performing better than amusics. A significant main effect of tone pair was also observed [$F(3,90) = 39.30$, $p < .001$]. Post-hoc pairwise *t*-tests (two-sided) with the Holm correction suggested that discrimination of tone pair 23-25 was significantly worse than that of other tone pairs (all *p*s < .001), and discrimination of tone pair 21-22 was also worse than that of 25-55 ($p < .001$). The group × tone pair interaction [$F(3,90) = 0.69$, $p = .561$] was not significant.

Correlation analyses indicated that tone discrimination performance was positively correlated between the speech and non-speech conditions for both amusics [$r(62) = .71$, $p < .001$] and controls [$r(62) = .65$, $p < .001$], suggesting a possible link in performing a lexical and non-lexical task.

B. Lexical tone production

The ten F_0 data points of each tone were converted into log scores, and a by-speaker *z*-score normalization was applied, in order to reduce inter-speaker variability. Fig. 3 shows the mean time-normalized F_0 contours (in *z*-score normalized log F_0) of the six Cantonese tones produced by the two groups. In order to account for the time-varying dynamics of these tones, a linear mixed effects model with group (amusic, control), tone (21, 22, 23, 33, 25, 55), and time (1-10) as fixed effects and participant (n = 32) as random effects were fit on the *z*-score normalized log F_0 values of each tone, averaged across the three repetitions by each participant. Results revealed significant main effects of tone [$F(5, 1866) = 1749.41$, $p < .001$] and time [$F(1, 1866) = 340.66$, $p < .001$], and a significant tone × time interaction [$F(5,1866) = 360.31$, $p < .001$]. There was no significant effect of group [$F(1, 30) = 0.0002$, $p = .988$], group × tone [$F(5, 1866) = 2.09$, $p = .064$], group × time [$F(1,1866) = 0.16$, $p = .688$], or group × tone × time interaction [$F(5, 1866) = 1.76$, $p = .117$]. This suggests that amusics produced their native tones with comparable F_0 trajectories as controls.

[Insert Fig. 3 about here]

Fig. 4 shows tone production accuracy (in percentage of correct responses) of the two groups as judged by eight native listeners. A repeated-measures ANOVA on rationalized arcsine transformed scores with factors of group (amusic, control) and tone (21, 22, 23, 33, 25, 55) revealed a significant main effect of tone [$F(5,150) = 22.14$, $p < .001$], but no group effect [$F(1,30) = 0.01$, $p = .913$] or group × tone interaction [$F(5,150) = 0.30$, $p = .910$]. Post-hoc pairwise *t*-tests (two-sided) with the Holm correction indicated that the low-falling tone 21 and the high-level tone 55 led to the highest recognition rates, followed by the two rising tones 23 and 25; the low-level and mid-level tones 22 and 33 gave rise to the worst recognition (all ps < .05).

[Insert Fig. 4 about here]

Confusion matrices of the eight native listeners' identification of the six Cantonese tones produced by the amusics and controls are shown in Table 2. The overall identification accuracy for amusics' tone production was 0.5803, which was similar to that for controls (0.5816). The two groups' tone production confusion matrices also exhibited similar values of Kappa statistic (amusics: 0.4964; controls: 0.4979), which is a measure of agreement between predicted and observed categories (Kuhn, 2008). This suggests that amusics' and controls' tone production led to similar confusion patterns for native listeners.

[Insert Table 2 about here]

C. Singing of "Happy birthday"

Fig. 5 shows singing ratings of the amusic and control groups by the eight Cantonese judges. A repeated-measures ANOVA with factors of group (amusic, control) and condition (rating by intuition versus guidelines) revealed significant main effects of group [$F(1,30) = 7.92$, $p = .009$] and condition [$F(1,30) = 53.81$, $p < .001$], but no group × condition interaction [$F(1,30) = 0.08$, $p = .776$]. The amusic group received significantly lower ratings on singing than the control group by judges based on both intuition and guidelines. Both groups received higher ratings when judges followed the rating guidelines than when using intuition.

[Insert Fig. 5 about here]

Fig. 6 shows boxplots of acoustic measurements in singing of “Happy birthday” by both groups. Repeated measures ANOVAs indicated that, compared to controls, amusics showed much larger pitch interval deviations [$F(1,30) = 5.30$, $p = .029$] and made significantly more numbers of pitch interval errors [$F(1,30) = 6.79$, $p = .014$]. Both groups made comparable numbers of contour errors [$F(1,30) = 0.87$, $p = .358$], while the group difference in time errors showed marginal significance [$F(1,30) = 3.85$, $p = .059$].

[Insert Fig. 6 about here]

In order to examine the patterns of pitch interval deviations by the two groups, a signed pitch interval deviation (in st) was calculated for each produced interval (in absolute value) as the difference from the corresponding notated interval (in absolute value). Thus, negative deviations indicate interval compressions, and positive deviations suggest interval expansions. Fig. 7 shows the mean (and SE) signed pitch interval deviations of the two groups against the notated/expected intervals for the song. A linear mixed-effects model was fit on the signed interval deviations, with group and notated interval as fixed effects and participants as random effects. The notated interval 0 st was excluded from the model, since theoretically only interval expansions are possible for this interval (Dalla Bella et al., 2009). There was a significant effect of notated interval on signed pitch interval deviation [$F(1,254) = 40.10$, $p < .001$)]: the bigger the notated interval, the greater the interval compression. Although no significant effect of group was found [$F(1,30) = 1.37$, $p = .251$], there was a marginally significant group × notated interval interaction [$F(1,254) = 3.44$, $p = .065$], owing to the fact that group differences tended to be larger for relatively large notated intervals (8 and 12 st).

[Insert Fig. 7 about here]

Correlation analyses suggested that, across all participants, singing ratings (averaged across intuition and guideline conditions) were negatively correlated with pitch interval deviations [$r(30) = -.75$, $p < .001$], numbers of pitch interval errors [$r(30) = -.81$, $p < .001$], and numbers of contour errors [$r(30) = -.44$, $p = .012$] in acoustic measurements. That is, as expected, better singing ratings were associated with smaller pitch interval deviations and fewer pitch interval and contour errors. There was no correlation between singing ratings and numbers of time errors [$r(30) = -.24$, $p = .178$]. This indicates that native listeners' singing ratings are related to or based on singers' pitch-related attributes in the current data.

D. Pitch thresholds

Fig. 8 shows pitch direction identification thresholds for speech syllables and piano tones of the two groups (15 amusics and 14 controls). A repeated-measures ANOVA on log-transformed thresholds with factors of group (amusic, control) and stimulus type (speech syllable versus piano tone) revealed a significant main effect of group [$F(1,27) = 6.51$, $p = .017$], but no effect of stimulus type [$F(1,27) = 2.43$, $p = .131$], or group × stimulus type interaction [$F(1,27) = 0.14$, $p = .711$]. This outcome indicates that amusics had higher pitch direction identification thresholds than controls for both speech syllables and piano tones.

[Insert Fig. 8 about here]

E. Correlations between lexical, musical, and psychophysical tasks

Correlation analyses were conducted between participants' pitch thresholds (averaged across speech syllable and piano tone conditions), MBEA global scores, scores on the tone perception tasks (averaged across speech and non-speech conditions), tone production accuracy (rationalized arcsine transformed percent-correct scores), and singing ratings (averaged across intuition and guideline conditions) for each group. For amusics, significant correlations were found between pitch thresholds and tone perception scores [Fig. 9A; $r(13) = -.67$, $p = .006$; the lower the pitch thresholds, the better the tone perception], and between MBEA global scores and singing ratings [Fig. 9C; $r(14) = .59$, $p = .016$; the higher the MBEA scores, the higher the singing ratings]. For controls, significant correlations were found between MBEA global and tone perception scores [Fig. 9B; $r(14) = .60$, $p = .014$]. No significant correlation was found between tone perception, tone production (Fig. 9D), and singing ratings for either group (all ps > 0.05).

[Insert Fig. 9 about here]

IV. DISCUSSION

This study investigated pitch processing in speech and music in individuals with congenital amusia who are native speakers of a complex tone language, Cantonese. The rich inventory of Cantonese tones afforded us the opportunity to reveal the real-world consequences of a music perception disorder on lexical tone perception and production of one's native language, singing, and psychophysical pitch thresholds in the population of Cantonese speakers, who have previously been shown superior pitch processing for speech and music compared to English and Mandarin speakers (Bidelman et al., 2013; Zheng et al., 2010, 2014). Our results indicate that, despite demonstrating impaired lexical tone perception, Cantonese amusics showed normal production of native tones with subtle pitch height and contour differences. By contrast, Cantonese amusics showed impairments in singing, producing larger pitch interval deviations, and more pitch interval errors than controls. Cantonese amusics also demonstrated higher pitch direction identification thresholds relative to controls for both speech syllables and piano tones. Significant correlations were found between pitch thresholds and speech perception, music perception and production, but not between speech perception and production for Cantonese amusics. These results are consistent with previous findings, providing further evidence that congenital amusia is domain-general language-independent pitch-processing deficit that is associated with severely impaired music perception and production, mildly impaired speech perception, and largely intact speech production.

Previous studies led to mixed results on amusics' pitch processing in speech versus non-speech conditions. In some studies, amusics showed normal performance on intonation perception in the speech condition but impaired performance on non-speech analogues (Ayotte et al., 2002; Liu et al., 2012a; Patel et al., 2005). In other studies, amusics were equally impaired for both types of stimuli (Jiang et al., 2010; Liu et al., 2010, 2012b). In the current study, Cantonese amusics showed impaired lexical tone discrimination in both speech and non-speech contexts. Consistent with this result, our recent study also indicated that Cantonese-speaking amusics were more likely than controls to confuse between acoustically similar tones in a lexical tone identification task (Liu et al., 2015b). It is worth noting that Cantonese is currently going through a sound change process, in which tones with similar acoustic features, e.g., tone pair 23-25 (two rising tones), 22-33 (two level tones), and 21-22 (two low tones), are merging into one single category (Law et al., 2013; Mok and Zuo, 2012; Mok et al., 2013). It is possible that, with impaired pitch processing abilities, amusics are more easily affected by such a sound change than controls. Further investigations are required to explore this possibility.

Furthermore, another reason why amusics and controls performed more poorly for these acoustically similar tone pairs than for the tone pair 25-55 may be due to the absence of context for helping listeners to determine the speaker's tonal space in tone perception. External context plays a role in how well tones may be discriminated (DiCanio, 2012; Francis et al., 2006), but its role was not examined here. It will be interesting to examine whether and to what extent amusics' tone perception is influenced by sentential context in future studies.

In tone languages such as Cantonese where multiple tones share similar pitch patterns (tones 23 and 25; tones 55, 33, and 22; tones 22 and 21), the requirement for tone production precision is likely increased. However, consistent with previous findings on Mandarin amusics (Nan et al., 2010), the current lexical tone production data by Cantonese amusics provide a similar result pointing to intact production but impaired perception of lexical tones in amusia. This evidence suggests that perceptual deficits in fine-grained discrimination of Cantonese tones are dissociable from production accuracy of the same tones. Indeed, acoustic analysis of the tone production data and confusion matrices of native listeners' identification of these tones revealed no difference in tone production between amusics and controls, even for the tones (e.g., 22-33, 21-22) that were difficult to discriminate for amusics. This finding is consistent with previous findings of English amusics' normal pitch control in story reading and vowel production (Liu et al., 2010), Mandarin amusics' normal production of native tones (Nan et al., 2010), and French-speaking amusics' intact imitation of speech intonation and vocal control (Hutchins and Peretz, 2012, 2013). Nevertheless, when asked to imitate the exact pitch and rhythm patterns of spoken utterances in Mandarin, Mandarin amusics still exhibited impaired performance relative to controls in terms of both pitch and rhythm matching (Liu et al., 2013). Together, these results suggest that amusics are proficient speakers of their native languages, without demonstrating obvious speech production problems.

Compared with English- and French-speaking amusics, fewer Cantonese-speaking amusics reported to have problems with singing (Peretz et al., 2008; Wong et al., 2012). The current data revealed that, despite the substantial overlap in singing proficiency between Cantonese amusics and controls (Fig. 5-7, Fig. 9C), amusics as a group still sang significantly worse than controls. Singing inaccuracy in Cantonese amusics mainly manifested in the compression of large pitch intervals (Fig. 7). This finding is consistent with interval compression by English and French amusics reported by Dalla Bella et al. (2009), although no effect of interval size was found for amusics of the same English and French background by Tremblay-Champoux et al. (2010). Furthermore, similar to the amusics in Tremblay-Champoux et al. (2010), the Cantonese amusics in the present study showed no impairment in rhythm processing in singing, whereas a few amusics demonstrated impaired rhythmic processing as indexed by tempo and temporal variability in Dalla Bella et al. (2009). Finally, while the current Cantonese amusics performed at the normal level on contour processing, English and French amusics made significantly more contour errors compared to controls (Dalla Bella et al., 2009; Tremblay-Champoux et al., 2010). Further studies are needed to examine whether the aforementioned differences in amusic singing are due to the different language background of these amusics (Cantonese versus English and French), or some other factors such as sampling error.

Similar to amusics from other language backgrounds (English, French, or Mandarin) (Foxton et al., 2004; Jiang et al., 2013; Liu et al., 2010, 2012a, 2012b; Tillmann et al., 2009), the Cantonese amusics as a group also demonstrated significantly higher thresholds for identification of pitch direction in speech and music in the present study, despite the findings that Cantonese speakers have cognitive and perceptual abilities comparable to English musicians for musical pitch processing (Bidelman et al., 2013). This result suggests that speaking Cantonese does not completely compensate for pitch-processing deficits in speech and music for individuals with congenital amusia.

Recent studies comparing amusics' performance on the perception versus the production of speech intonation, pitch direction, and lexical tones show that amusics may have preserved pitch production abilities, despite demonstrating impaired perception (Hutchins and Peretz, 2012; Liu et al., 2010; Loui et al., 2008; Nan et al., 2010). Dual-route models of perception and production have been proposed to account for the disconnection between pitch production and perception in amusia and in other similar studies (Griffiths, 2008; Hickok and Poeppel, 2007; Hutchins and Moreno, 2013; Loui, 2015). In particular, these models suggest that two different, independent pathways are responsible for the production versus perception of vocal information in speech and music, with the dorsal stream integrating sensorimotor commands while the ventral stream analyzing category-based representations (Griffiths, 2008; Hickok and Poeppel, 2007; Hutchins and Moreno, 2013; Loui, 2015). Consequently, it is possible to observe the dissociation between pitch production and perception abilities in amusia due to the different processing streams involved. The current findings of retained lexical tone production but impaired lexical tone perception in Cantonese amusics provide further evidence for dual-route models of perception and production. However, the significant correlation between amusics' musical perception and singing abilities seems to suggest that singing ability is, at least to some extent, constrained by musical perception ability. This conclusion is likely due to the higher demand for precision in pitch production in music than in speech (Liu et al., 2013). The relationship between perception acuity and production accuracy in speech and music warrants further investigations (Hutchins and Moreno, 2013).

V. CONCLUSION

In summary, this study examined pitch perception and production in speech and music in Cantonese-speaking individuals with congenital amusia, a neurodevelopmental disorder of musical processing. Results indicate a significant impairment in lexical tone perception, but not in production for these amusics. Demonstrating impaired singing, these amusics also exhibited elevated pitch direction identification thresholds for both speech and music. While a significant correlation was observed between music perception and singing abilities of these amusics, a lack of correlation was found between their speech perception and production abilities. These findings provide further evidence that congenital amusia is domain-general language-independent pitch-processing deficit that is associated with severely impaired music perception and production, mildly impaired speech perception, and largely intact speech production.

ACKNOWLEDGEMENTS

This work was supported by National Science Foundation Grant BCS-1125144, the Liu Che Woo Institute of Innovative Medicine at The Chinese University of Hong Kong, the US National Institutes of Health grant R01DC013315, the Research Grants Council of Hong Kong grants 477513 and 14117514, and the Stanley Ho Medical Development Foundation to P.C.M.W, grants (M4080396) from College of Humanities, Arts & Social Sciences at Nanyang Technological University, M4011393 and MOE2015-T2-1-120 from the Ministry of Education Singapore to A.H.D.C., and grants from the Natural Sciences and Engineering Research Council of Canada, Canada Institute of Health Research and a Canada Research Chair to I.P. The authors would like to thank Hilda Chan, Antonio Cheung, Shanshan Hu, Judy Kwan, Doris Lau, Joseph Lau, Vina Law, Wen Xu, and Kynthia Yip for their assistance in this research.

References

Abramson, A. S. (**1962**). *The vowels and tones of Standard Thai: Acoustical measurements and experiments*, Indiana University Research Center in Anthropology, Folklore and Linguistics, Pub. 20, Bloomington: Indiana University.

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., Daligault, S., et al. (**2013**). “Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex,” Brain, **136**, 1639–1661. doi:10.1093/brain/awt082

Anderson, S., Himonides, E., Wise, K., Welch, G., and Stewart, L. (**2012**). “Is there potential for learning in amusia? A study of the effect of singing intervention in congenital amusia,” Ann. N. Y. Acad. Sci., **1252**, 345–353. doi:10.1111/j.1749-6632.2011.06404.x

Ayotte, J., Peretz, I., and Hyde, K. L. (**2002**). “Congenital amusia: A group study of adults afflicted with a music-specific disorder,” Brain J. Neurol., **125**, 238–251.

Bent, T., Bradlow, A. R., and Wright, B. A. (**2006**). “The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds,” J. Exp. Psychol. Hum. Percept. Perform., **32**, 97–103. doi:10.1037/0096-1523.32.1.97

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (**2011**). “Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem,” J. Cogn. Neurosci., **23**, 425–434. doi:10.1162/jocn.2009.21362

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (**2011**). "Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch," Brain Cogn., **77**, 1–10. doi:10.1016/j.bandc.2011.07.006

Bidelman, G. M., Hutka, S., and Moreno, S. (**2013**). "Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and Music," PLoS ONE, **8**, e60676. doi:10.1371/journal.pone.0060676

Boersma, P., and Weenink, D. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int., **5**, 341–345.

Chao, Y. R. (**1930**). "A system of tone letters," Maître Phon., **45**, 24–27.

Ciocca, V., and Lui, J. Y. K. (**2003**). "The development of the perception of Cantonese lexical tones," J. Multiling. Commun. Disord., **1**, 141–147. doi:10.1080/1476967031000090971

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (**2007**). "Singing proficiency in the general population," J. Acoust. Soc. Am., **121**, 1182–1189.

Dalla Bella, S., Giguère, J.-F., and Peretz, I. (**2009**). "Singing in congenital amusia," J. Acoust. Soc. Am., **126**, 414–424. doi:10.1121/1.3132504

Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (**2009**). "Absolute pitch among students in an American music conservatory: association with tone language fluency," J. Acoust. Soc. Am., **125**, 2398–2403. doi:10.1121/1.3081389

DiCanio, C. T. (**2012**). "Cross-linguistic perception of Itunyoso Trique tone," J. Phon., **40**, 672–688. doi:10.1016/j.wocn.2012.05.003

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., and Spector, T. (**2001**). "Genetic correlates of musical pitch recognition in humans," Science, **291**, 1969–1972. doi:10.1126/science.291.5510.1969

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (**2004**). "Characterization of deficits in pitch perception underlying 'tone deafness,'" Brain, **127**, 801–810. doi:10.1093/brain/awh105

Francart, T., van Wieringen, A., and Wouters, J. (**2008**). "APEX 3: a multi-purpose test platform for auditory psychophysical experiments," J. Neurosci. Methods, **172**, 283–293. doi:10.1016/j.jneumeth.2008.04.020

Francis, A. L., and Ciocca, V. (**2003**). "Stimulus presentation order and the perception of lexical tones in Cantonese," J. Acoust. Soc. Am., **114**, 1611–1621.

Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., and Chu, P. C. Y. (**2006**). "Extrinsic context affects perceptual normalization of lexical tone," J. Acoust. Soc. Am., **119**, 1712–1726.

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., and Wicha, N. Y. Y. (**2011**). "Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change," Front. Audit. Cogn. Neurosci., **2**, 146. doi:10.3389/fpsyg.2011.00146

Griffiths, T. D. (**2008**). "Sensory systems: auditory action streams?," Curr. Biol. CB, **18**, R387-388. doi:10.1016/j.cub.2008.03.007

Henry, M. J., and McAuley, J. D. (**2010**). "On the prevalence of congenital amusia," Music Percept., **27**, 413–418. doi:10.1525/mp.2010.27.5.413

Henry, M. J., and McAuley, J. D. (**2013**). "Failure to apply signal detection theory to the Montreal Battery of Evaluation of Amusia may misdiagnose amusia," Music Percept. Interdiscip. J., **30**, 480–496. doi:10.1525/mp.2013.30.5.480

Hickok, G., and Poeppel, D. (**2007**). "The cortical organization of speech processing," Nat. Rev. Neurosci., **8**, 393–402. doi:10.1038/nrn2113

Holm, S. (**1979**). "A simple sequentially rejective multiple test procedure," Scand. J. Stat., **6**, 65–70. doi:10.2307/4615733

Hutchins, S., Gosselin, N., and Peretz, I. (**2010**). "Identification of Changes along a Continuum of Speech Intonation is Impaired in Congenital Amusia," Front. Psychol., **1**, 236. doi:10.3389/fpsyg.2010.00236

Hutchins, S., and Moreno, S. (**2013**). "The Linked Dual Representation model of vocal perception and production," Front. Psychol., **4**, 825. doi:10.3389/fpsyg.2013.00825

Hutchins, S., and Peretz, I. (**2012**). "Amusics can imitate what they cannot discriminate," Brain Lang., **123**, 234–239. doi:10.1016/j.bandl.2012.09.011

Hutchins, S., and Peretz, I. (**2013**). "Vocal pitch shift in congenital amusia (pitch deafness)," Brain Lang., **125**, 106–117. doi:10.1016/j.bandl.2013.01.011

Hutchins, S., Zarate, J. M., Zatorre, R. J., and Peretz, I. (**2010**). "An acoustical study of vocal pitch matching in congenital amusia," J. Acoust. Soc. Am., **127**, 504–512. doi:10.1121/1.3270391

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., and Peretz, I. (**2007**). "Cortical thickness in congenital amusia: when less is better than more," J. Neurosci., **27**, 13028–13032. doi:10.1523/JNEUROSCI.3039-07.2007

Hyde, K. L., and Peretz, I. (**2004**). “Brains that are out of tune but in time,” Psychol. Sci., **15**, 356–360. doi:10.1111/j.0956-7976.2004.00683.x

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., and Peretz, I. (**2006**). “Morphometry of the amusic brain: a two-site study,” Brain, **129**, 2562–2570. doi:10.1093/brain/awl204

Hyde, K. L., Zatorre, R. J., and Peretz, I. (**2011**). “Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia,” Cereb. Cortex, **21**, 292–299. doi:10.1093/cercor/bhq094

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., Chen, X., and Yang, Y. (**2012**). “Amusia results in abnormal brain activity following inappropriate intonation during speech comprehension,” PLoS ONE, **7**, e41411. doi:10.1371/journal.pone.0041411

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (**2010**). “Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese,” Neuropsychologia, **48**, 2630–2639. doi:10.1016/j.neuropsychologia.2010.05.009

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (**2011**). “Fine-grained pitch discrimination in congenital amusics with Mandarin Chinese,” Music Percept., **28**, 519–526.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., and Yang, Y. (**2012**). “Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics,” Mem. Cognit., **40**, 1109–1121. doi:10.3758/s13421-012-0208-2

Jiang, C., Lim, V. K., Wang, H., and Hamm, J. P. (**2013**). "Difficulties with pitch discrimination influences pitch memory performance: Evidence from congenital amusia," PLoS ONE, **8**, e79216. doi:10.1371/journal.pone.0079216

Jones, J. L., Lucker, J., Zalewski, C., Brewer, C., and Drayna, D. (**2009**). "Phonological processing in adults with deficits in musical pitch recognition," J. Commun. Disord., **42**, 226–234. doi:10.1016/j.jcomdis.2009.01.001

Jones, J. L., Zalewski, C., Brewer, C., Lucker, J., and Drayna, D. (**2009**). "Widespread auditory deficits in tune deafness," Ear Hear., **30**, 63–72. doi:10.1097/AUD.0b013e31818ff95e

Kalmus, H., and Fry, D. B. (**1980**). "On tune deafness (dysmelodia): frequency, development, genetics and musical background," Ann. Hum. Genet., **43**, 369–382.

Kuhn, M. (**2008**). "Building predictive models in R using the caret package," J. Stat. Softw., **28**, 1–26.

Law, S.-P., Fung, R., and Kung, C. (**2013**). "An ERP Study of Good Production vis-à-vis Poor Perception of Tones in Cantonese: Implications for Top-Down Speech Processing," PLoS ONE, **8**, e54396. doi:10.1371/journal.pone.0054396

Lee, C.-Y., and Hung, T.-H. (**2008**). "Identification of Mandarin tones by English-speaking musicians and nonmusicians," J. Acoust. Soc. Am., **124**, 3235–3248. doi:10.1121/1.2990713

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am., **49**, 467–477.

Lin, H. B., and Repp, B. H. (**1989**). "Cues to the perception of Taiwanese tones," Lang. Speech, **32 (Pt 1)**, 25–44.

Liu, F., Jiang, C., Pfordresher, P. Q., Mantell, J. T., Xu, Y., Yang, Y., and Stewart, L. (**2013**). "Individuals with congenital amusia imitate pitches more accurately in singing than in speaking: Implications for music and language processing," Atten. Percept. Psychophys., **75**, 1783–1798. doi:10.3758/s13414-013-0506-1

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., and Stewart, L. (**2012**). "The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers," PloS One, **7**, e30374. doi:10.1371/journal.pone.0030374

Liu, F., Jiang, C., Wang, B., Xu, Y., and Patel, A. D. (**2015**). "A music perception disorder (congenital amusia) influences speech comprehension," Neuropsychologia, **66**, 111–118. doi:10.1016/j.neuropsychologia.2014.11.001

Liu, F., Maggu, A. R., Lau, J. C. Y., and Wong, P. C. M. (**2015**). "Brainstem encoding of speech and musical stimuli in congenital amusia: Evidence from Cantonese speakers," Front. Hum. Neurosci., **8**, 1029. doi:10.3389/fnhum.2014.01029

Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (**2010**). "Intonation processing in congenital amusia: Discrimination, identification and imitation," Brain, **133**, 1682–1693. doi:10.1093/brain/awq089

Liu, F., Xu, Y., Patel, A. D., Francart, T., and Jiang, C. (**2012**). "Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers," Brain Cogn., **79**, 209–215. doi:10.1016/j.bandc.2012.03.008

Liu, S., and Samuel, A. G. (**2004**). “Perception of Mandarin lexical tones when F0 information is neutralized,” Lang. Speech, **47**, 109–138.

Loui, P. (**2015**). “A Dual-Stream Neuroanatomy of Singing,” Music Percept., **32**, 232–241. doi:10.1525/mp.2015.32.3.232

Loui, P., Alsop, D., and Schlaug, G. (**2009**). “Tone deafness: A new disconnection syndrome?,” J. Neurosci., **29**, 10215–10220. doi:10.1523/JNEUROSCI.1701-09.2009

Loui, P., Guenther, F. H., Mathys, C., and Schlaug, G. (**2008**). “Action-perception mismatch in tone-deafness,” Curr. Biol. CB, **18**, R331–R332. doi:10.1016/j.cub.2008.02.045

Loui, P., Kroog, K., Zuk, J., and Schlaug, G. (**2011**). “Relating pitch awareness to phonemic awareness in children: implications for tone-deafness and dyslexia,” Front. Audit. Cogn. Neurosci., **2**, 111. doi:10.3389/fpsyg.2011.00111

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User’s Guide*, Lawrence Erlbaum Associates, 526 pages.

Mok, P. K. P., and Zuo, D. (**2012**). “The separation between music and speech: Evidence from the perception of Cantonese tones,” J. Acoust. Soc. Am., **132**, 2711–2720. doi:10.1121/1.4747010

Mok, P. P. K., Zuo, D., and Wong, P. W. Y. (**2013**). “Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese,” Lang. Var. Change, **25**, 341–370. doi:10.1017/S0954394513000161

Moreau, P., Jolicoeur, P., and Peretz, I. (**2009**). "Automatic brain responses to pitch changes in congenital amusia," Ann. N. Y. Acad. Sci., **1169**, 191–194. doi:10.1111/j.1749-6632.2009.04775.x

Moreau, P., Jolicœur, P., and Peretz, I. (**2013**). "Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials," Brain Cogn., **81**, 337–344. doi:10.1016/j.bandc.2013.01.004

Nan, Y., Sun, Y., and Peretz, I. (**2010**). "Congenital amusia in speakers of a tone language: Association with lexical tone agnosia," Brain, **133**, 2635–2642. doi:10.1093/brain/awq178

Oldfield, R. C. (**1971**). "The assessment and analysis of handedness: The Edinburgh inventory," Neuropsychologia, **9**, 97–113.

Patel, A. D., Foxton, J. M., and Griffiths, T. D. (**2005**). "Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech," Brain Cogn., **59**, 310–313. doi:10.1016/j.bandc.2004.10.003

Peretz, I. (**2008**). "Musical Disorders: From Behavior to Genes," Curr. Dir. Psychol. Sci., **17**, 329–333. doi:10.1111/j.1467-8721.2008.00600.x

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., and Jutras, B. (**2002**). "Congenital amusia: a disorder of fine-grained pitch discrimination," Neuron, **33**, 185–191.

Peretz, I., Brattico, E., Järvenpää, M., and Tervaniemi, M. (**2009**). "The amusic brain: In tune, out of key, and unaware," Brain, **132**, 1277–1286. doi:10.1093/brain/awp055

Peretz, I., Champod, A. S., and Hyde, K. L. (**2003**). "Varieties of musical disorders The Montreal Battery of Evaluation of Amusia," Ann. N. Y. Acad. Sci., **999**, 58–75.

Peretz, I., Cummings, S., and Dubé, M.-P. (**2007**). "The genetics of congenital amusia (tone deafness): A family-aggregation study," Am. J. Hum. Genet., **81**, 582–588. doi:10.1086/521337

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L., Gagnon, B., Trimmer, C., Paquette, S., et al. (**2008**). "On-line identification of congenital amusia," Music Percept., **25**, 331–343.

Peretz, I., Nguyen, S., and Cummings, S. (**2011**). "Tone language fluency impairs pitch discrimination," Front. Audit. Cogn. Neurosci., **2**, 145. doi:10.3389/fpsyg.2011.00145

Pfordresher, P. Q., and Brown, S. (**2009**). "Enhanced production and perception of musical pitch in tone language speakers," Atten. Percept. Psychophys., **71**, 1385–1398. doi:10.3758/APP.71.6.1385

R Core Team (**2014**). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/*, R Foundation for Statistical Computing, Vienna, Austria.

Schellenberg, E. G., and Trehub, S. E. (**2008**). "Is there an Asian advantage for pitch memory?," Music Percept., **25**, 241–252. doi:10.1525/mp.2008.25.3.241

So, C. K., and Best, C. T. (**2010**). "Cross-language perception of non-native tonal contrasts: effects of native phonological and phonetic influences," Lang. Speech, **53**, 273–293.

Stagray, J. R., and Downs, D. (**1993**). “Differential sensitivity for frequency among speakers of a tone and a nontone language,” J. Chin. Linguist., **21**, 143–63.

Studebaker, G. A. (**1985**). “A ‘rationalized’ arcsine transform,” J. Speech Hear. Res., **28**, 455–462.

Thompson, W. F., Marin, M. M., and Stewart, L. (**2012**). “Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis,” Proc. Natl. Acad. Sci., **109**, 19027–19032. doi:10.1073/pnas.1210344109

Tillmann, B., Nguyen, S., Grimault, N., Gosselin, N., and Peretz, I. (**2011**). “Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages,” Front. Audit. Cogn. Neurosci., **2**, 120. doi:10.3389/fpsyg.2011.00120

Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., and Peretz, I. (**2011**). “Fine-grained pitch processing of music and speech in congenital amusia,” J. Acoust. Soc. Am., **130**, 4089–4096. doi:10.1121/1.3658447

Tillmann, B., Schulze, K., and Foxton, J. M. (**2009**). “Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds,” Brain Cogn., **71**, 259–264. doi:10.1016/j.bandc.2009.08.003

Tong, X., Lee, S. M. K., Lee, M. M. L., and Burnham, D. (**2015**). “A Tale of Two Features: Perception of Cantonese Lexical Tone and English Lexical Stress in Cantonese-English Bilinguals,” PLOS ONE, **10**, e0142896. doi:10.1371/journal.pone.0142896

Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M.-A., and Peretz, I. (**2010**). "Singing proficiency in congenital amusia: imitation helps," Cogn. Neuropsychol., **27**, 463–476. doi:10.1080/02643294.2011.567258

Varley, R., and So, L. K. H. (**1995**). "Age effects in tonal comprehension in Cantonese," J. Chin. Linguist., **23**, 76–97.

Vuvan, D. T., Nunes-Silva, M., and Peretz, I. (**2015**). "Meta-analytic evidence for the non-modularity of pitch processing in congenital amusia," Cortex, **69**, 186–200. doi:10.1016/j.cortex.2015.05.002

Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica, **49**, 25–47.

Wise, K. J., and Sloboda, J. A. (**2008**). "Establishing an empirical profile of self-defined 'tone deafness': Perception, singing performance and self-assessment," Retrieved from https://jyx.jyu.fi/dspace/handle/123456789/19385. Retrieved from https://jyx.jyu.fi/dspace/handle/123456789/19385

Wong, A. M.-Y., Ciocca, V., and Yung, S. (**2009**). "The Perception of Lexical Tone Contrasts in Cantonese Children With and Without Specific Language Impairment (SLI)," J. Speech Lang. Hear. Res., **52**, 1493–1509. doi:10.1044/1092-4388(2009/08-0170)

Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L.-H., and Peretz, I. (**2012**). "Effects of culture on musical pitch perception," PloS One, **7**, e33424. doi:10.1371/journal.pone.0033424

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (**2007**). "Musical experience shapes human brainstem encoding of linguistic pitch patterns," Nat. Neurosci., **10**, 420–422. doi:10.1038/nn1872

Xu, Y. (**2013**). "ProsodyPro — A tool for large-scale systematic prosody analysis," Proc. Tools Resour. Anal. Speech Prosody, Aix-en-Provence, France, 7–10. Presented at the TRASP 2013.

Yip, M. (**2002**). *Tone*, Cambridge University Press, Cambridge, UK, 380 pages.

Zheng, H.-Y., Minett, J. W., Peng, G., and Wang, W. S.-Y. (**2010**). "The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study," Lang. Cogn. Process., **27**, 184–209. doi:10.1080/01690965.2010.520493

Zheng, H.-Y., Peng, G., Chen, J.-Y., Zhang, C., Minett, J. W., and Wang, W. S.-Y. (**2014**). "The Influence of Tone Inventory on ERP without Focal Attention: A Cross-Language Study," Comput. Math. Methods Med., **2014**, e961563. doi:10.1155/2014/961563

Figure captions

Fig. 1. Time-normalized F_0 contours (in Hertz, or Hz) of the four tone pairs used in the perception tasks: (A) bei22 [鼻, "nose"] - bei33 [臂, "arm"]; (B) jyu23 [雨, "rain"] - jyu25 [鱼, "fish"]; (C) min21 [綿, "cotton"] - min22 [麵, "noodle"]; (D) tong25 [糖, "sweet (candy)"] - tong55 [湯, "soup"]. Note: 22, 33, 23, 25, 21, and 55 correspond to the Cantonese low-level (Tone 6), mid-level (Tone 3), low-rising (Tone 5), high-rising (Tone 2), low-falling (Tone 4), and high-level (Tone 1) tones, respectively.

Fig. 2. Performance (in *d'*) of the 16 amusics and 16 controls in the tone perception tasks: (A) speech condition, and (B) non-speech condition. Note: Tone21_22 stands for the stimulus pairs containing Tone21 and/or Tone22 either in "same" (Tone21-Tone21; Tone22-Tone22) or "different" (Tone21-Tone22; Tone22-Tone21) conditions, and the same is for Tone22_33, Tone23_25, and Tone25_55.

Fig. 3. Mean time-normalized F_0 contours (in *z*-score normalized log F_0) of the six Cantonese tones (A-F) produced by amusics and controls, averaged across 16 participants in each group. Note: Tone55, Tone25, Tone33, Tone21, Tone23, and Tone22 correspond to the Cantonese high-level (Tone 1), high-rising (Tone 2), mid-level (Tone 3), low-falling (Tone 4), low-rising (Tone 5), and low-level (Tone 6) tones, respectively. Grey error bars reflect standard error.

Fig. 4. Tone production accuracy (in percentage of correct responses) of 16 amusics and 16 controls as judged by 8 native listeners. Note: Tone21, Tone22, Tone23, Tone25, Tone33, and Tone55 correspond to the Cantonese low-falling (Tone 4),

low-level (Tone 6), low-rising (Tone 5), high-rising (Tone 2), mid-level (Tone 3), and high-level (Tone 1) tones, respectively.

Fig. 5. Boxplots of singing ratings of 16 amusics and 16 controls as evaluated by 8 Cantonese listeners by provided guidelines (A) and based on intuition (B). The accuracy rating guidelines (1-8) were developed by Wise and Sloboda (Anderson et al., 2012; Wise and Sloboda, 2008). Ratings by intuition were made on the scale of 1-8, with 8 being very in-tune. These boxplots contain the extreme of the lower whisker, the lower hinge, the median, the upper hinge, and the extreme of the upper whisker. The two hinges are the first and third quartile, and the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box.

Fig. 6. Boxplots of acoustic measurements in singing of "Happy birthday" by 16 amusics and 16 controls: (A) pitch interval deviation (in semitones), (B) number of pitch interval errors (out of the total 23/24 intervals), (C) number of contour errors (out of the total 23/24 contours), and (D) number of time errors (out of the total 24/25 notes). The black dots denote individual participants in each group, with those at the same horizontal level having identical values. The data points that lie beyond the extremes of the whiskers are outliers, which are further denoted by small open circles.

Fig. 7. Mean (and SE) signed pitch interval deviations (in semitones) of amusics (grey dashed lines) and controls (black straight lines) against notated intervals (in semitones) in singing of "Happy birthday".

Fig. 8. Pitch direction identification thresholds (in semitones) of 15 amusics and 14 controls for piano tones (A) and speech syllables (B).

Fig. 9. Scatter plots of participants' scores across different tasks: (A) pitch thresholds (averaged across speech syllable and piano tone conditions; in semitones) against scores on the tone perception tasks (averaged across speech and non-speech conditions; in d'), (B) MBEA global scores (in percentage of correct responses out of the total 180 trials in the MBEA tasks) against scores on the tone perception tasks, (C) MBEA global scores against singing ratings (averaged across intuition and guideline conditions), and (D) tone production accuracy (rationalized arcsine transformed percent-correct scores) against scores on the tone perception tasks.

Table 1. Characteristics of the amusic (*n* = 16, 12 female, 4 male, all right-handed) and control (*n* = 16, 12 female, 4 male, all right-handed) groups.

Group	Control mean (SD)	Amusic mean (SD)	*t*	*p*
Age	24.94 (9.18)	25.19 (10.51)	-0.07	0.943
Education	16.00 (3.03)	14.31 (2.02)	1.85	0.074
Musical training	1.13 (1.82)	0.94 (1.77)	0.30	0.770
Scale	27.88 (1.67)	21.75 (2.46)	8.24	< .001
Contour	27.88 (1.50)	21.88 (2.31)	8.72	< .001
Interval	28.00 (1.55)	20.19 (1.94)	12.59	< .001
Rhythm	29.25 (0.93)	23.00 (2.63)	8.95	< .001
Meter	27.63 (2.96)	20.50 (4.47)	5.31	< .001
Memory	29.13 (1.09)	26.94 (2.82)	2.90	0.007
Pitch composite	83.75 (2.96)	63.81 (3.90)	16.29	< .001
MBEA Global	94.31 (3.13)	74.58 (5.31)	12.81	< .001

Age, education, and musical training are in years; scores on the six MBEA subtests (scale, contour, interval, rhythm, meter, and memory) are in number of correct responses out of 30 (Peretz et al., 2003); the pitch composite score is the sum of the scale, contour, and interval scores; MBEA global score is the percentage of correct responses out of the total 180 trials; *t* is the statistic of the Welch two sample *t*-test (two-tailed, *df* = 30).

Table 2. Confusion matrices (in percentages) of Cantonese tones produced by 16 amusics and 16 controls, as judged by eight native listeners.

Amusic	Tone55	Tone25	Tone33	Tone21	Tone23	Tone22
Tone55	**73.18**	0.26	11.72	0.52	2.34	3.65
Tone25	0.00	**61.72**	0.52	0.52	27.08	0.78
Tone33	19.01	0.26	**32.55**	5.47	6.51	25.78
Tone21	0.26	0.00	3.13	**76.30**	0.00	21.09
Tone23	0.78	37.50	3.13	2.60	**58.07**	2.34
Tone22	6.77	0.26	48.96	14.58	5.99	**46.35**
Control	Tone55	Tone25	Tone33	Tone21	Tone23	Tone22
Tone55	**70.05**	0.00	17.45	1.82	1.56	10.16
Tone25	0.52	**65.10**	0.78	0.26	32.03	0.78
Tone33	18.23	0.26	**32.81**	9.38	2.34	25.78
Tone21	0.26	0.00	4.69	**76.04**	0.78	17.97
Tone23	0.00	34.38	5.21	0.52	**62.76**	3.13
Tone22	10.94	0.26	39.06	11.98	0.52	**42.19**

Tones in columns are produced/expected categories, and those in rows are perceived categories.

Figure 1

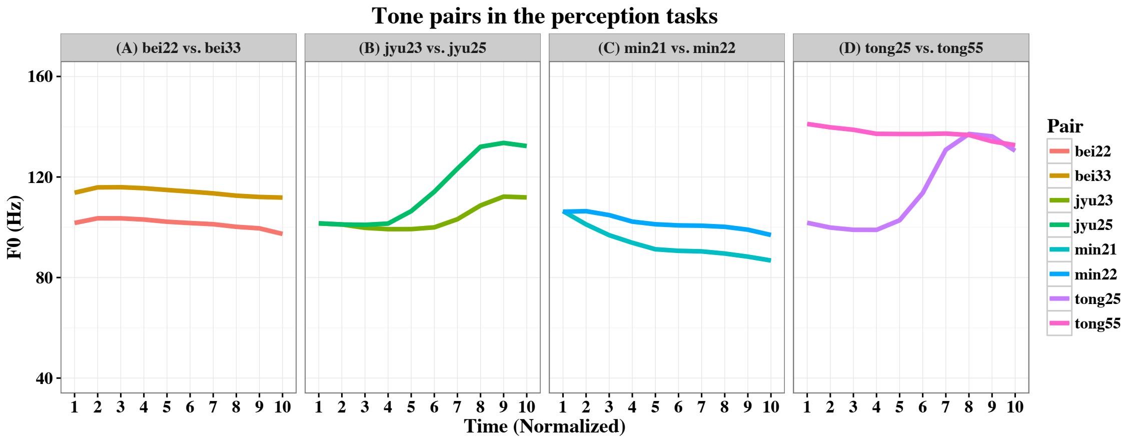

Tone pairs in the perception tasks
(A) bei22 vs. bei33
(B) jyu23 vs. jyu25
(C) min21 vs. min22
(D) tong25 vs. tong55
F0 (Hz)
160
120
80
40
1 2 3 4 5 6 7 8 9 10
Time (Normalized)
Pair
bei22
bei33
jyu23
jyu25
min21
min22
tong25
tong55

Figure 2

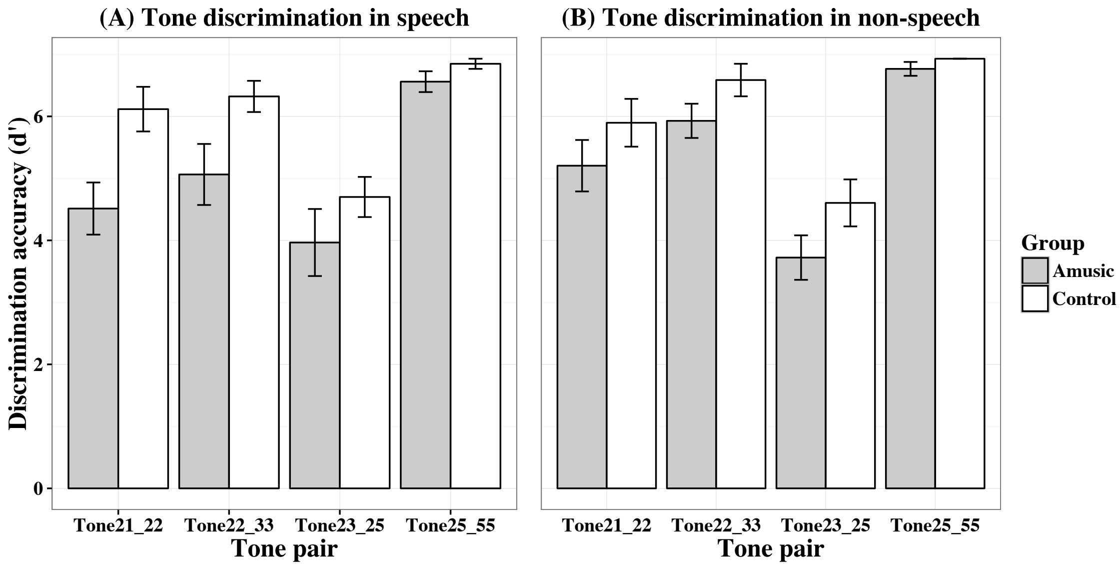
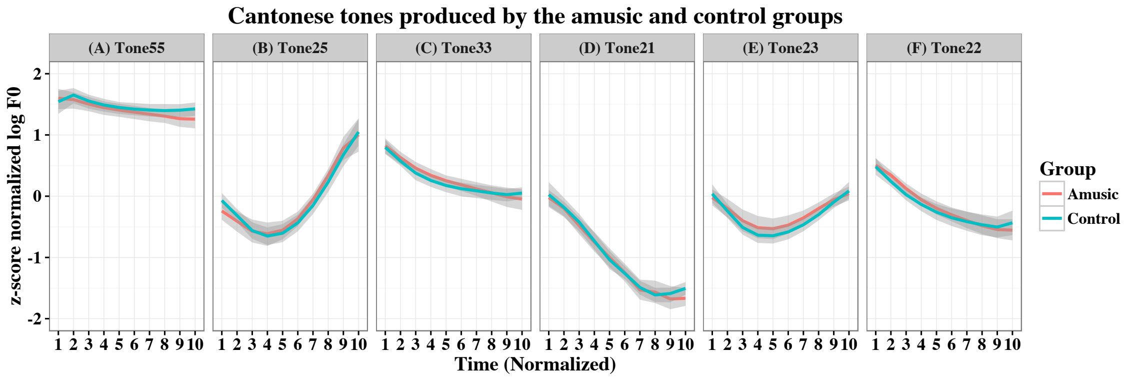

(A) Tone discrimination in speech
(B) Tone discrimination in non-speech
Discrimination accuracy (d')
0
2
4
6
Tone21_22
Tone22_33
Tone23_25
Tone25_55
Tone pair
Group
Amusic
Control

Figure 3



Cantonese tones produced by the amusic and control groups
(A) Tone55
(B) Tone25
(C) Tone33
(D) Tone21
(E) Tone23
(F) Tone22
z-score normalized log F0
2
1
0
-1
-2
1 2 3 4 5 6 7 8 9 10
Time (Normalized)
Group
Amusic
Control

Figure 4

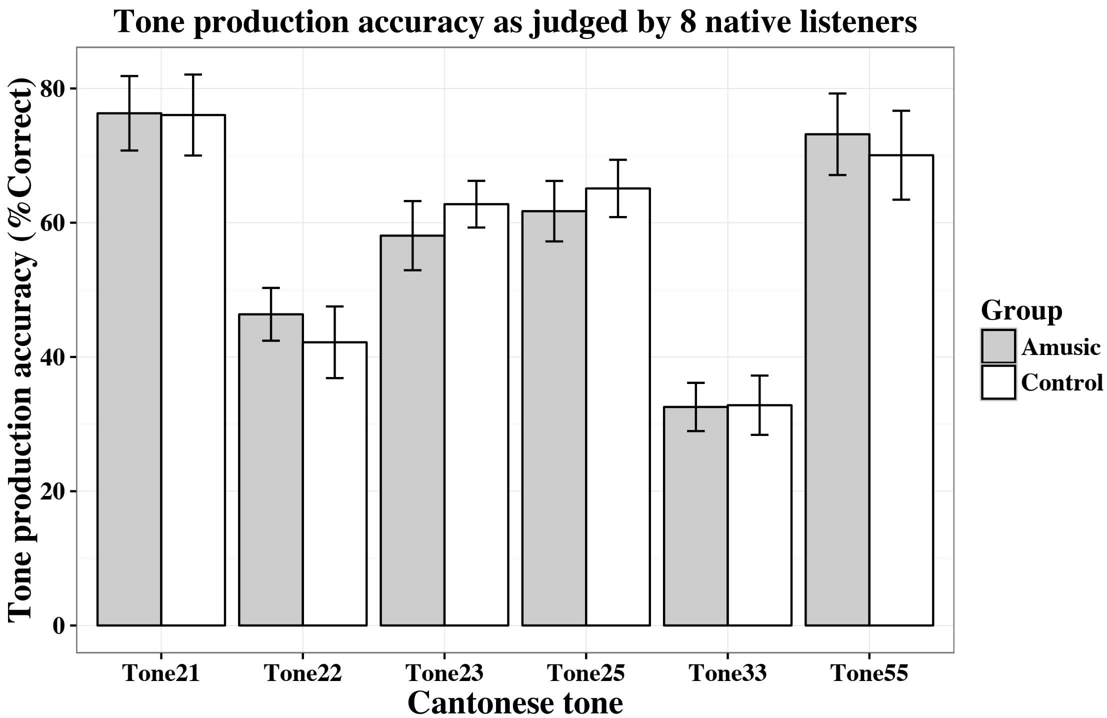

Tone production accuracy as judged by 8 native listeners
Tone production accuracy (%Correct)
80
60
40
20
0
Tone21
Tone22
Tone23
Tone25
Tone33
Tone55
Cantonese tone
Group
Amusic
Control

Figure 5

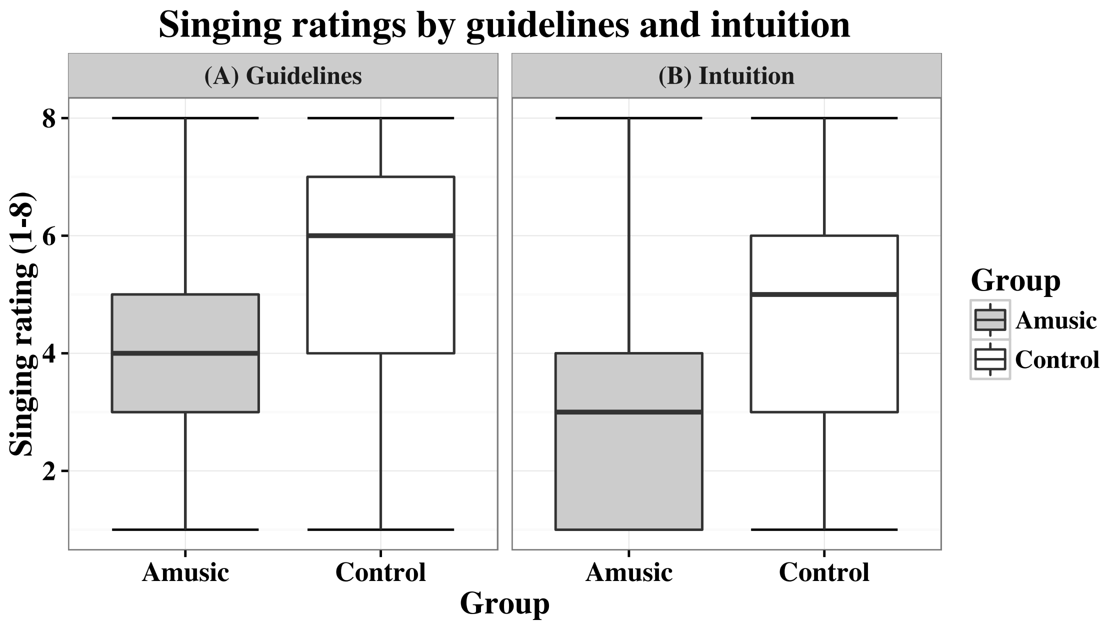
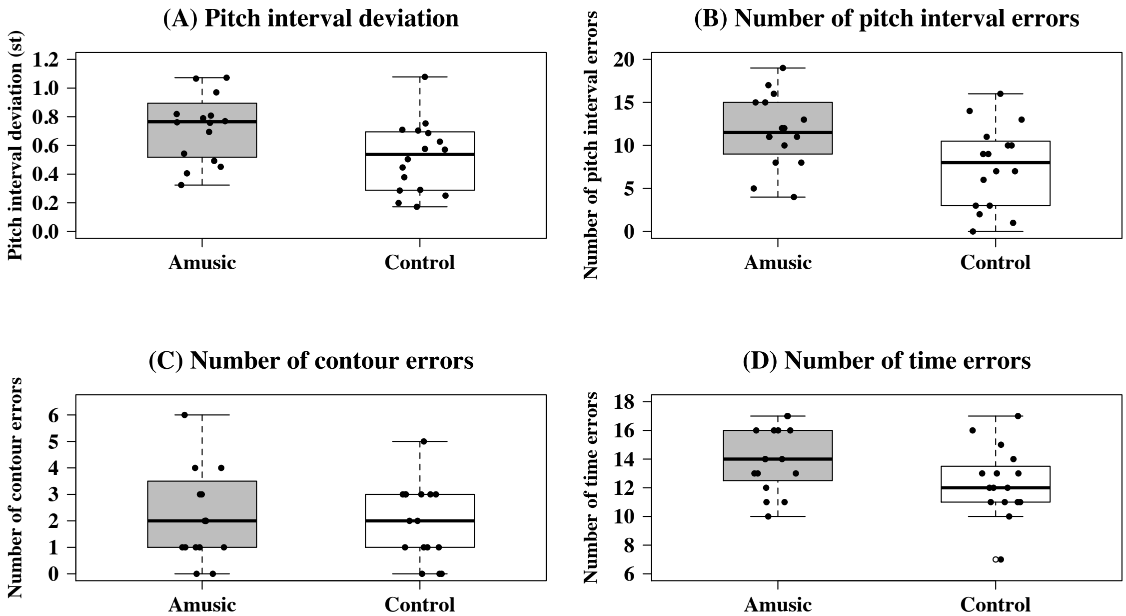

Singing ratings by guidelines and intuition
(A) Guidelines
(B) Intuition
Singing rating (1-8)
8
6
4
2
Amusic
Control
Amusic
Control
Group
Group
Amusic
Control

Figure 6



(A) Pitch interval deviation
Pitch interval deviation (st)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Amusic
Control
(B) Number of pitch interval errors
Number of pitch interval errors
20
15
10
5
0
Amusic
Control
(C) Number of contour errors
Number of contour errors
6
5
4
3
2
1
0
Amusic
Control
(D) Number of time errors
Number of time errors
18
16
14
12
10
8
6
Amusic
Control

Figure 7

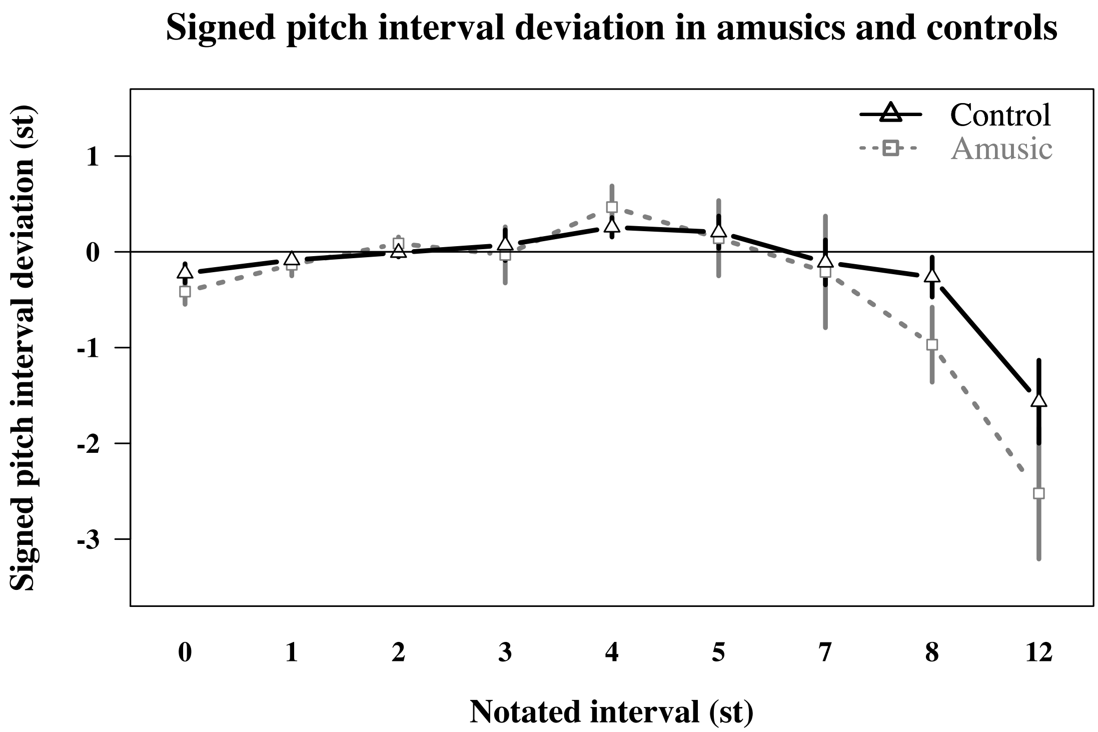
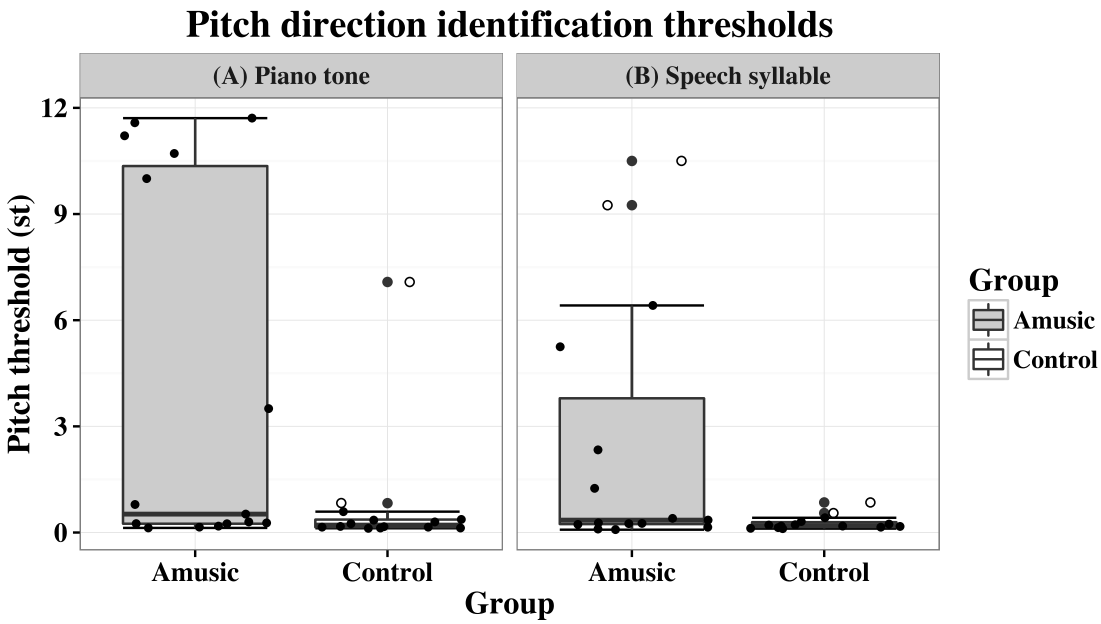
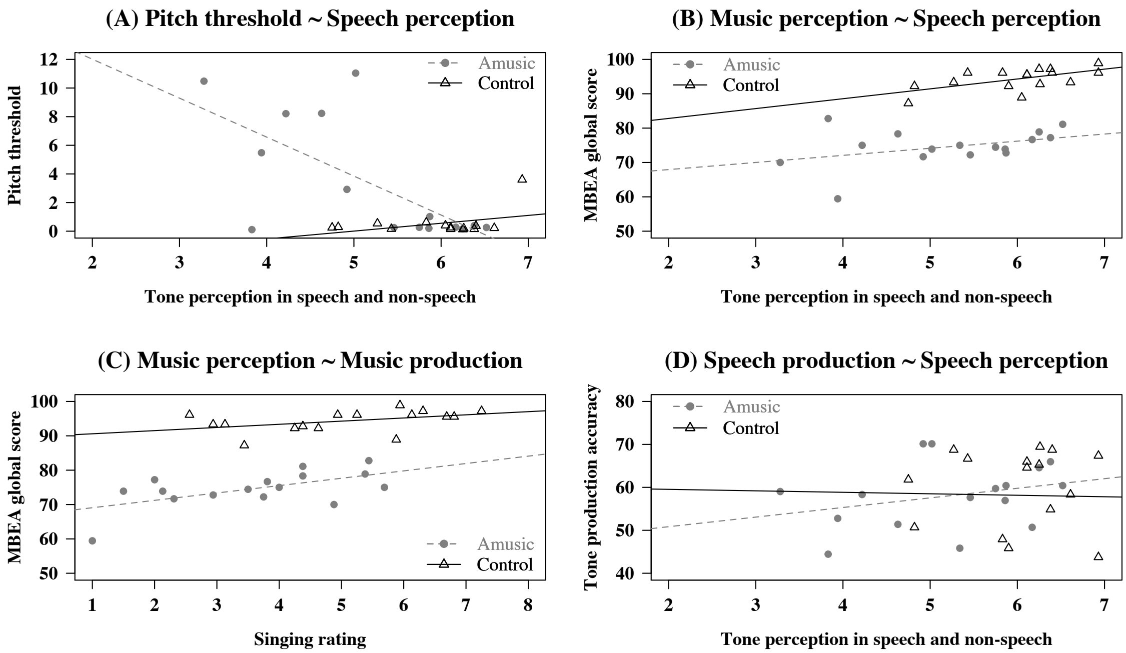

Signed pitch interval deviation in amusics and controls
Control
Amusic
1
0
-1
-2
-3
Signed pitch interval deviation (st)
0
1
2
3
4
5
7
8
12
Notated interval (st)

Figure 8



Pitch direction identification thresholds
(A) Piano tone
(B) Speech syllable
Pitch threshold (st)
12
9
6
3
0
Amusic
Control
Amusic
Control
Group
Group
Amusic
Control

Figure 9


(A) Pitch threshold ~ Speech perception
Pitch threshold
Tone perception in speech and non-speech
Amusic
Control
(B) Music perception ~ Speech perception
MBEA global score
Tone perception in speech and non-speech
Amusic
Control
(C) Music perception ~ Music production
MBEA global score
Singing rating
Amusic
Control
(D) Speech production ~ Speech perception
Tone production accuracy
Tone perception in speech and non-speech
Amusic
Control